UNIVERSITI TEKNOLOGI MARA

QUERY EXPANSION AND LAYERS FILTRATION IN TOURIST SPATIAL INFORMATION RETRIEVAL

ROSILAWATI BINTI ZAINOL

Thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Faculty of Computer and Mathematical Sciences

February 2010

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution for any other degree or qualification.

In the event that my thesis is found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree and agree to be subjected to the disciplinary rules and regulations of Universiti Teknologi MARA.

Name of Candidate	ROSILAWATI BINTI ZAINOL
Candidate's ID No.	2002100244
Programme	Doctor of Philosophy
Faculty	Faculty of Computer and Mathematical Sciences
Thesis Title	Query Expansion and Layers Filtration in Tourist Spatial Information Retrieval
Signature of Candidate	.....
Date	25 February 2010

ABSTRACT

Increasing demand in retrieving information with geographic reference has challenged scholars to conduct studies in producing efficient method in retrieving spatial data and evaluating web-based geographic information system (GIS) on the internet, also known as web-based GIS. Tourist relies on maps throughout his visits in order to fulfil his information needs. Online maps with geographic information system (GIS) capabilities that are able to cater for tourist's spatial needs are few on the internet. Therefore, this research's objectives are to examine online map preferences by Malaysians, to evaluate the current online GIS maps capabilities, and to discover an efficient method to retrieve spatial information required by tourist. In order to achieve these objectives, tourist's spatial information databases of Shah Alam City Centre, Kuala Lumpur Golden Triangle and Taiping Historical Town Centre are developed. Furthermore, since spatial data are stored in separate layers according to their features: point, polyline and polygon, gazetteer method is found to be the most suitable method of handling query of this data type. Therefore, exact match and partial match keyword queries are used. Query lists are collected from surveys. A survey on online map preference is carried out among 489 Malaysians in Klang Valley to determine their preferences in conducting online search on spatial information. This survey contributes to the collection of query lists and relevant judgment. In addition, relevant judgment is also derived from interviews. Four experiments were carried out using all test collections. First experiment examines query using GoogleMaps Malaysia. The second experiment examines query by using the original SuperWebGIS 2.1. The third experiment examines query using enhanced SuperWebGIS 2.1 by case insensitivity and layer filtration. Finally, the fourth experiment examines query using enhanced SuperWebGIS 2.1 by query expansion and the combination of the third technique. Results show that combination of query expansion, case insensitivity and layer filtration technique yields 70% better results than using case insensitivity and layer filtration alone and 98% better than without using any of the three. The combination of these experiments has produced a system named MuSafIR (**Mu**safir **S**patial **I**nformation **R**etrieval).

ACKNOWLEDGEMENT

Praise to Allah S.W.T. the All Mighty for showering me with good experiences throughout this study period and for all that has been bestowed on me. It is Allah's ascendancy that the study is completed.

This research would not have been possible without support of many people. I wish to express my heartiest gratitude to my supervisor Professor Dr. Zainab Abu Bakar who was abundantly helpful and offered invaluable assistance, support and guidance.

My sincere gratitude also goes to Associate Professor Dr. Sayed Jamaludin Sayed Ali for his co-supervision and guidance. I would like to acknowledge Professor Dr. Ishak Ab. Ghani for his assistance and guidance. Special thanks to my technical assistant, Cik Norhanizawati Zainol @ Zainoon and my proofreaders, Encik Halim Idris and Mr. Kanapathy Allimuthu.

I would like to take this opportunity to thank the Ministry of Higher Education for the financial assistance rendered through the research period, University of Malaya for awarding me the two and a half years study leave and Universiti Teknologi MARA for the extension of time given to me.

Last but not least I wish to express my love and gratitude to my husband, sons, Ahmad Hariz Halim and Ahmad Hafiz Halim and other family members for their understanding, support and endless love through the duration of my studies and to all my friends who have given me the motivation and moral support to complete this research.

TABLE OF CONTENT